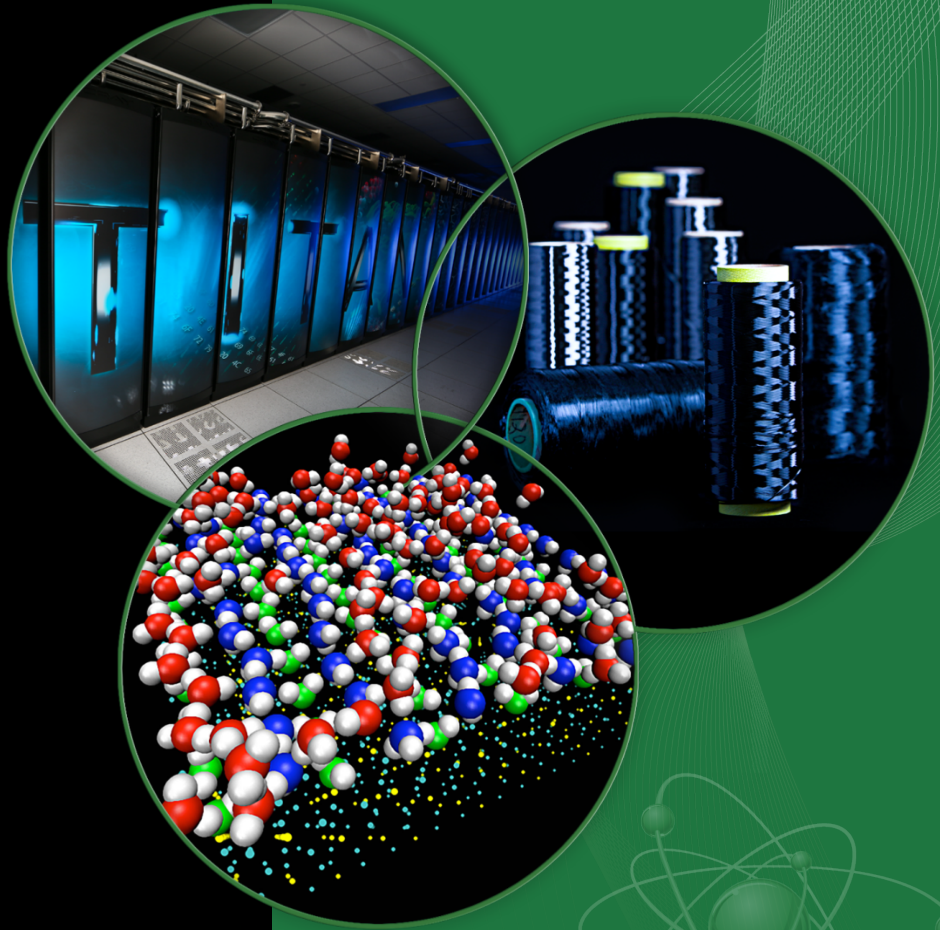
The OLCF "Data Group"

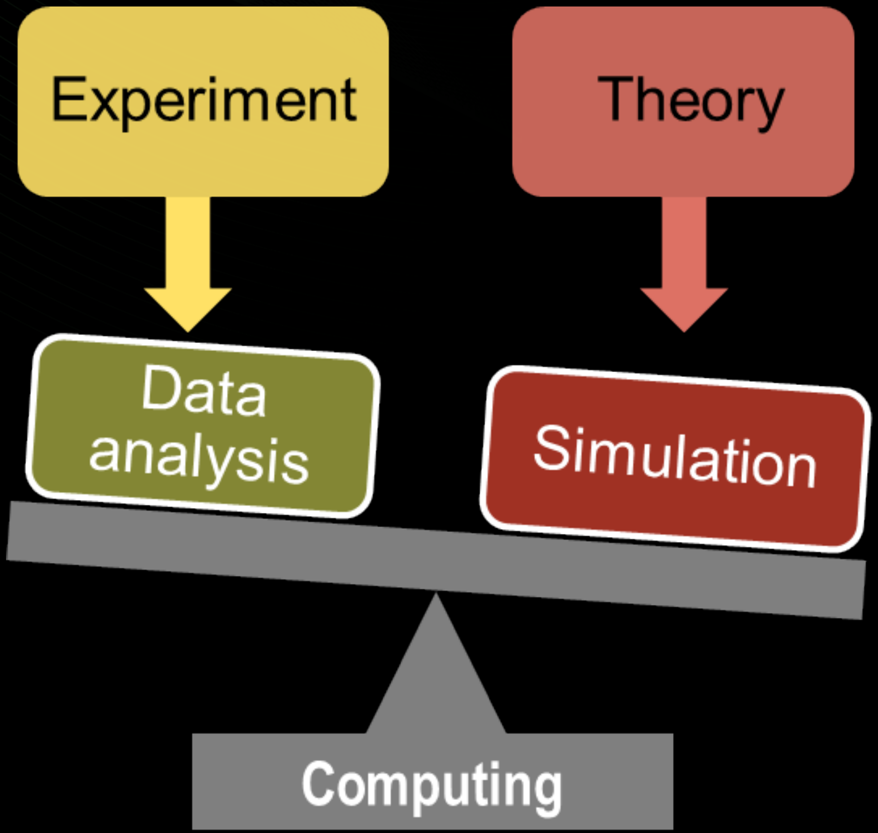
Sreenivas "Rangan" Sukumar

Data Scientist and Group Leader
National Center for Computational Sciences
Oak Ridge Leadership Computing Facility



ORNL is managed by UT-Battelle
for the US Department of Energy

Group Mission and Vision

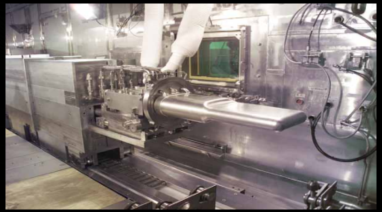


Mission:
Design and build creative solutions for data-driven discovery in science domains at scale and performance using diverse compute architectures.

Vision:
On-Demand Data, Analytics and Workflows

Group Future : Scientific Data Facility

Spallation Neutron Source

Center for Nanophase Material Sciences


Manufacturing Demonstration Facility


+

OLCF
OAK RIDGE LEADERSHIP COMPUTING FACILITY

CADES

=

Facilities of the future

The "Data" Group : People

Admin: Jessica West

Data Science Liaisons
Rangan Sukumar
John Harney
Valentine Anantharaj
Scott Klasky (M)
George Ostrouchov (M)

Visualization Liaisons
Mike Matheson
Jamison Daniel
Benjamin Hernandez-Arreguin
Kat Engstrom
David Pugmire (M)

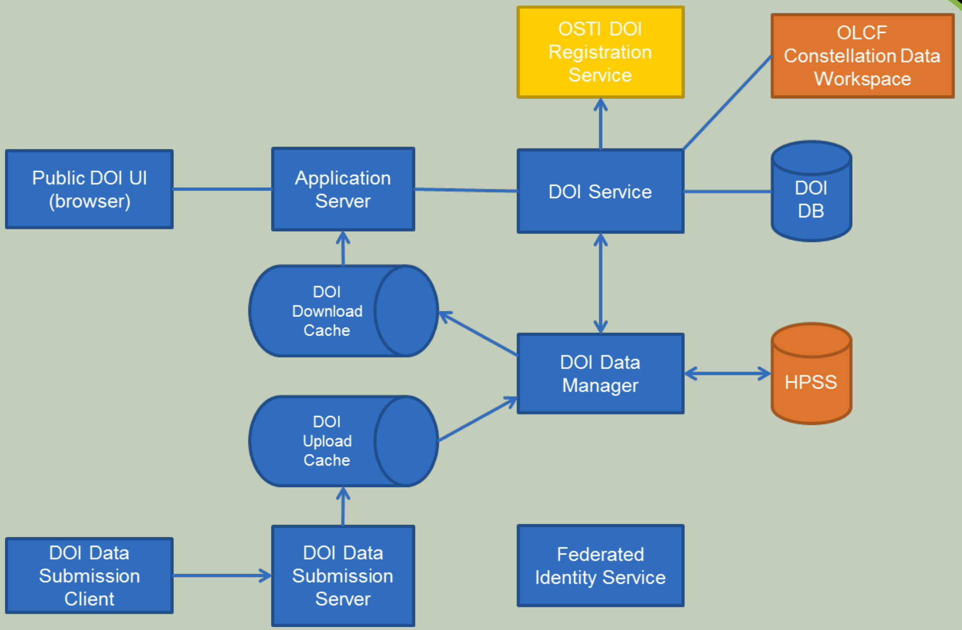
Production/Software
Dale Stansberry
Brian Smith
Norbert Podhroszki (M)

We need your help.

Survey Link:

http://goo.gl/QCit1z

Roadmap: On-Demand Data



DOIs

2016 Status: Pilot

Work with Sudharshan Vazhkudai

Data Portals

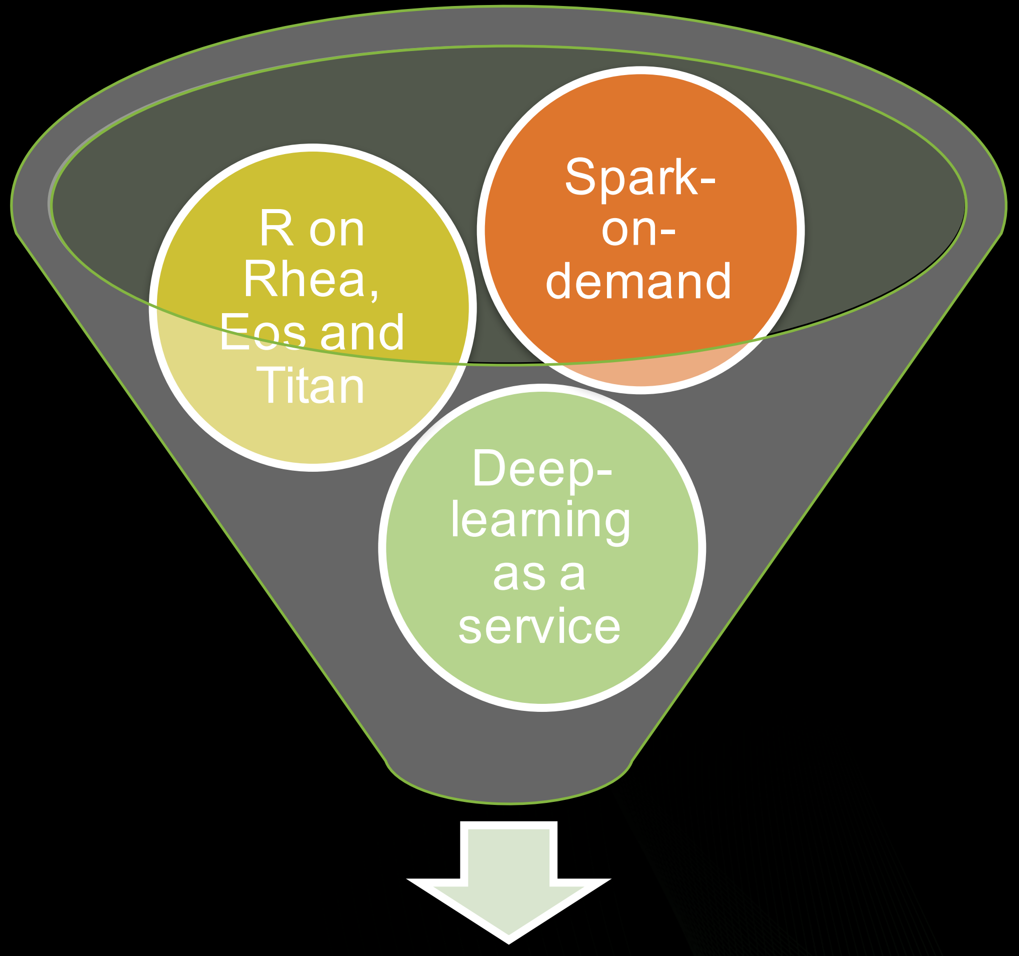
Peta-byte scale "Github" for Scientific Data

Collaboration with Technology Integration Group @ OLCF, OSTI and CADES

Roadmap: On-Demand Analytics

2016: Mini-Apps for Big Data

Kernel	Mathematical Form	Algorithms
Least Squares	$\sum XX^T$	PCA, ICA, Naïve Bayes, Linear Regression, Logistic Regression
Convolution	$\int f(s)a^{st}dt$	Fourier, wavelet, z-transforms, deep learning, blind separation, image reconstruction
Distance	$\lvert X_i - X_j \rvert^k$	Covariance matrix, ray-tracing, k-means, k-NN
Matrix Decomposition	$\min_{w\geq 0, H\geq 0} \parallel X - WH \parallel_F^2$	Recommender systems, Spatiotemporal data mining, signal processing
Optimization	$X^{t+1} \rightarrow \alpha PX^t + (1-\alpha)g$	Label propagation, iterative optimization
Sequence processing	$lev_{a,b}(i,j) = \begin{cases} \max(i,j) \quad if \min(i,j) \\ min \begin{cases} lev_{a,b}(i-1,j) \\ lev_{a,b}(i,j-1) \\ lev_{a,b}(i-1,j-1 \end{cases} \end{cases}$	Approximate text-search, sequence alignment problems
Graph-theoretic	$det(QAQ^T)$	N-ary search, pattern extraction



Architecture-Agnostic Analytic App-Store

Roadmap: On-Demand Workflows

2016: Testing Today's Tools

Pegasus, Kepler, FireWorks, Apache Taverna, Galaxy, PANDA

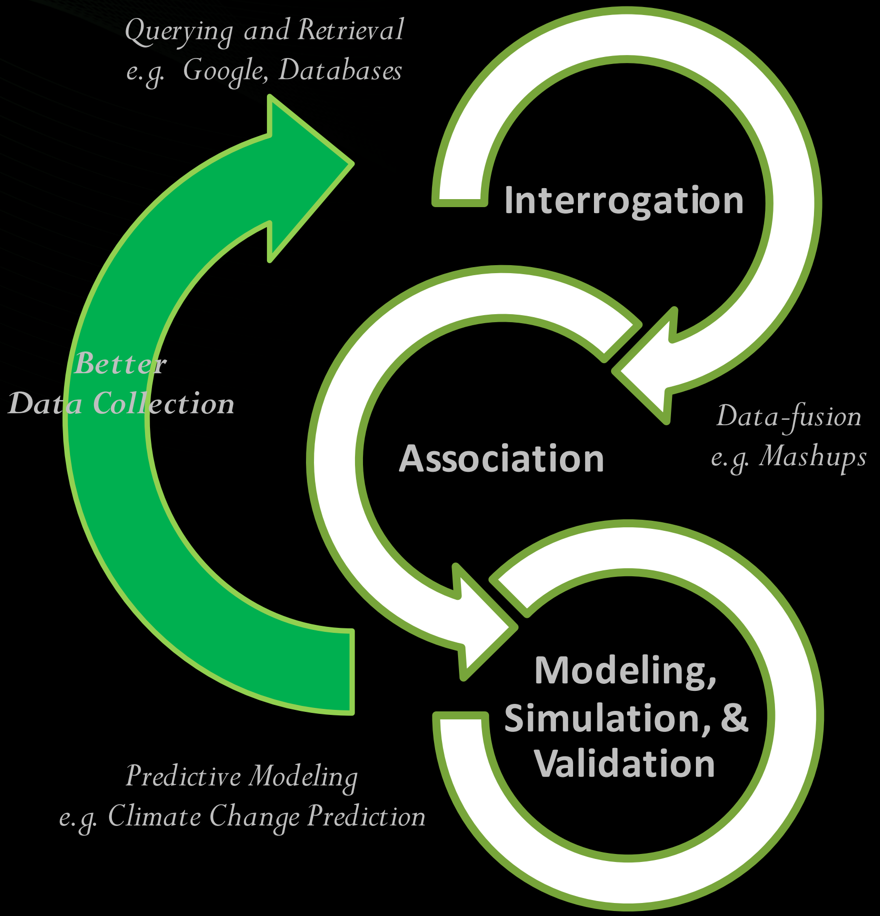
Eos 2.0 ?

Workflow Infrastructure			
Federated ID Management		Certificate Management	
Pilot Job Launcher	Resource Manager	Data-Transfer	Workflow State Tracker
HPC	Streaming (FPGA)	Globus endpoints	Graph

OAK RIDGE National Laboratory

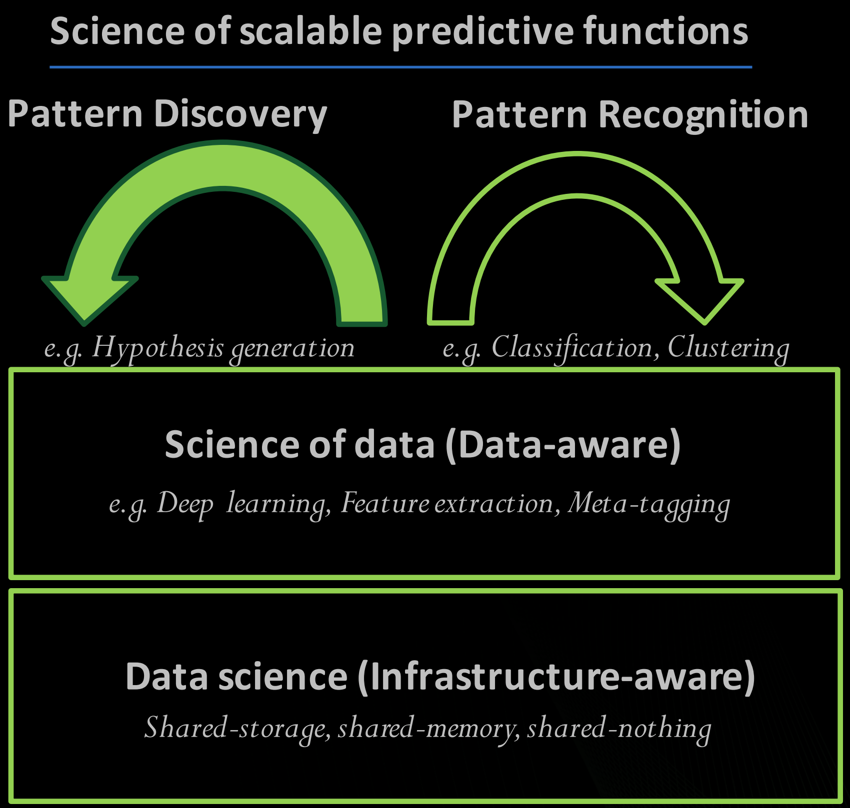
What have we started doing ? Data Science Support

The Lifecycle of Data-Driven Discovery



Domain Scientist's View

The Process of Data-Driven Discovery



Data Scientist's View

Building Knowledge Discovery Ecosystems

OAK RIDGE National Laboratory

What have we started doing ? Data Science Support

Analytic services
- Data mining
- Semantic analysis
- Data fusion

Data services
- Data transfer tools
- Metadata harvesting/ management
- Indexing, discovery, dissemination

Simulation services
- Simulation frame-works
- Scalable debuggers
- Scientific libraries

System software and middleware services
- MPI
- ADIOS
- Map reduce
- HIVE
- Key value stores
- Graph databases
- SQL databases
- Message queues
- . . .

Infrastructure services
- HPC compute
- Utility compute
- Advanced networking
- Parallel file systems
- Network storage
- Archival storage
- Object storage
- . . .

Integrate with Compute and Data Environment for Science

What are we learning ? Science vs. Industry

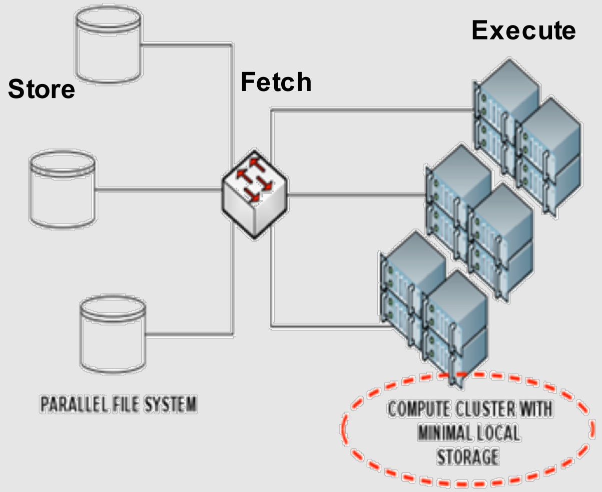
Big Data	Science	Industry
Data (Structured)	Vector, Matrix, Tensor	Table, Key-Values, Objects
Data (Unstructured)	Mesh, Images (Physics-based)	Documents, Images (Camera)
Visualization	Voxel, Surface, Point Clouds	Word Cloud, Parallel Co-ordinates, BI Tools
Validation	Cross-validation (ROC curves, statistical significance)	Manual / Subject matter expert, A/B testing
Extract, Transform, Load	Fourier, Wavelet, Laplace, etc. Cartesian, Radial, Toroidal, etc.	File-format transformations e.g. CSV to VRML
Search (Query)	Properties such as periodicity, self-similarity, anomaly, etc.	SQL, SPARQL, etc. (Sum, Average, Groupby)
Funding Model	Non-profit grand challenge (Answer matters)	Value-driven (Cost matters)

What are we learning ? : HPC vs. Big Data

HPC: Forward Problem

$$F(u) = \int_{-\infty}^{\infty} f(x) e^{-i2\pi ux} dx$$

$$HPC : f(x) = \sin(x)$$



- Data analysis algorithms are designed for functionality. Scaling and performance is an afterthought.
- Speed-up and scale-up are a function of architecture, data characteristics and algorithmic-workflow (as viewed from the Amdahl's law perspective).

Big Data: Inverse Problem

Latency Expectation	Real-time (with interactivity)
	Batch (response time not critical)
Access Pattern	Random (unpredictable access)
	Sequential (list-like access)
	Permutation (data is moved re-distributed often)
Working Set	All
	Partial
	Iterative
Data Type	Structured (tables, matrices)
	Unstructured (text, binary files)
	Media (images, video)
i/o profile	Read-heavy (loading data to memory)
	Write-heavy (large intermediate result sets)
Complexity	Low (data access with small compute operations)
	Medium
	High (data and compute intensive operations)

Different algorithms have different workload patterns

What are we learning ?

- Either, invest in customized hardware that are expensive ($1-10 M) and will NOT allow popular open-source/commercial software tools.

- Or, implement scalable theory-inspired algorithms on commodity or HPC clusters which in general takes a lot of effort without performance guarantees.

Massively parallel processing databases

PostgreSQL, Greenplum, Neo4j, mongoDB

Distributed Analytics on Storage

hadoop, mahout, Spark, Spark MLlib

Distributed-memory Analytics

MATLAB, R, sas, Knowledge Discovery Toolbox

"Analytics is retrieval"

"Take compute to cheap storage"

"Algorithm is made to work on distributed memory-chunks"

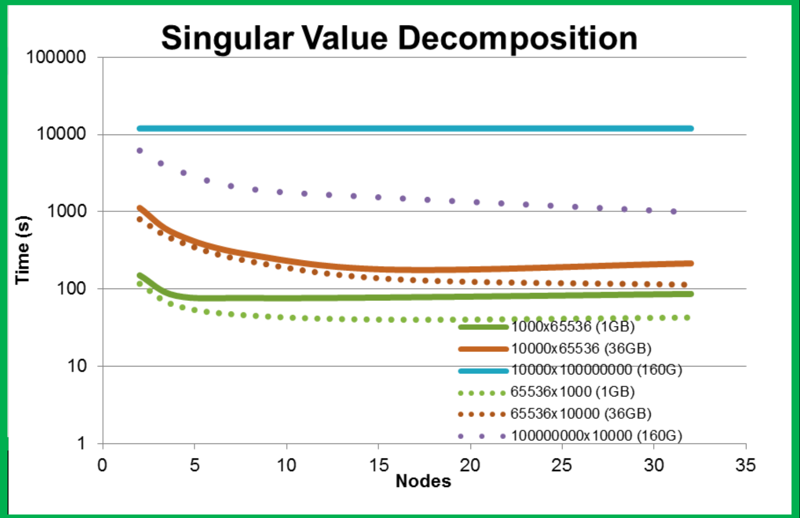
How are we solving some of these challenges ?

Apache Spark on Demand

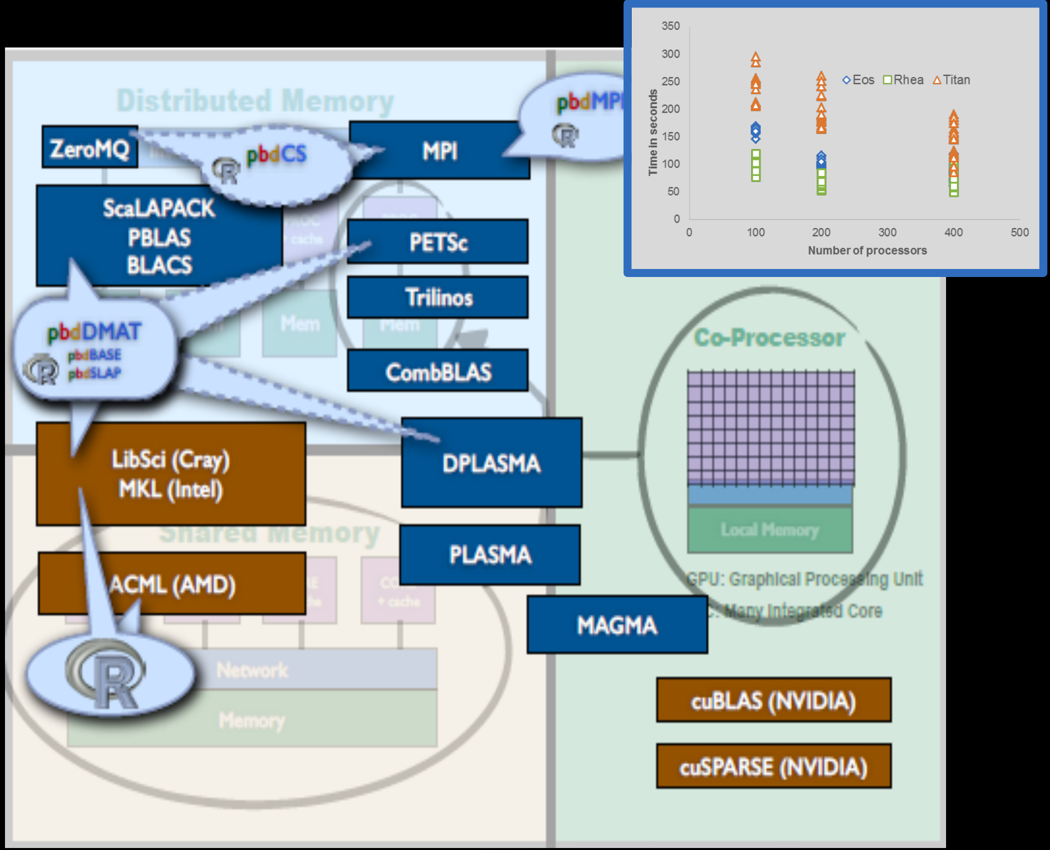
```
object SVD {
  def main(args: Array[String]) {
    if (args.length != 1) {
      System.err.println("Usage: SVD <input>")
      System.exit(1)
    }

    // Spark context
    val conf = new SparkConf().setAppName("SVD")
    val sc = new SparkContext(conf)

    // Load and parse the data file.
    val rows = sc.textFile(args(0)).map { line =>
      val values = line.split(' ').map(_.toDouble)
      Vectors.dense(values)
    }
    rows.cache()
    val mat = new RowMatrix(rows)

    // Compute SVD.
    val svd = mat.computeSVD(50)

    // Relevant Output
    val singular = svd.s
    val u = svd.U
    val v = svd.V

    sc.stop()
  }
}
```

Work with Seung-Hwan Lim

Spark now runs on Rhea



R on Rhea, Eos and Titan



Work with George Ostrouchov

Tutorial and demos presented on Day 0 sessions and will be available on the OLCF website

Oak Ridge National Laboratory

What have we started doing ? Visualization Support

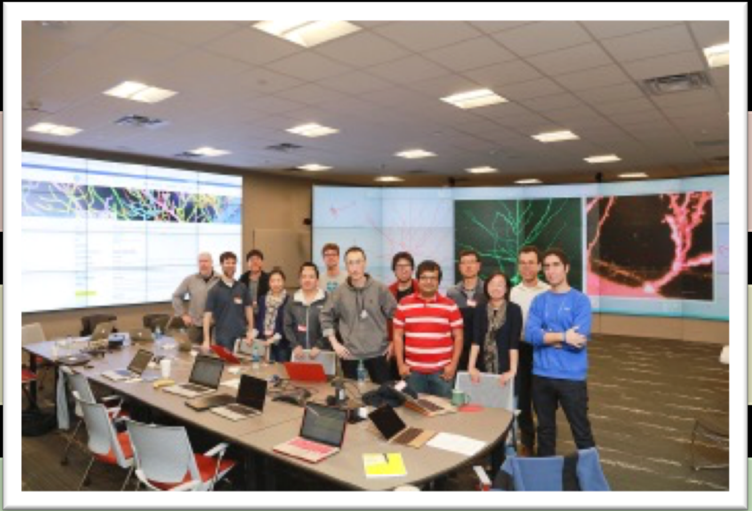
General Visualization Tools: VisIt, Paraview, Ensight, NICE DCV

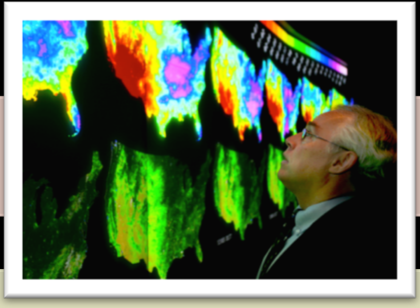
Domain Visualization Tools: VMD, vaa3d, vapor

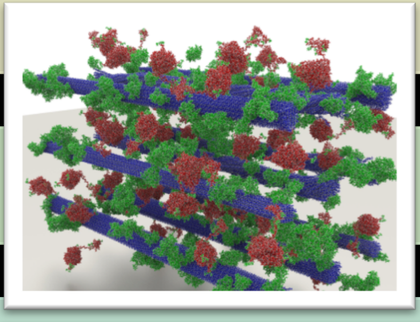
Libraries: ADIOS, OpenGL, OptiX

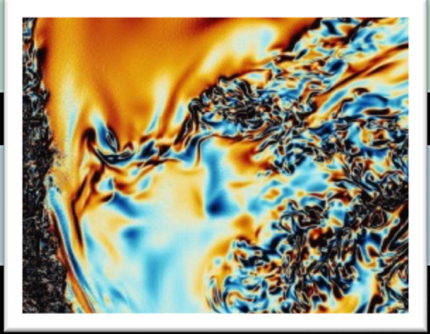
Rendering Tools: Maya, Unity, Blender, Custom

Hardware: Everest, Titan GPUs









What are we learning ?

Oculus Rift

Personalized Visualization

Remote Visualization

ParaView

NICE

DCV desktop cloud visualization

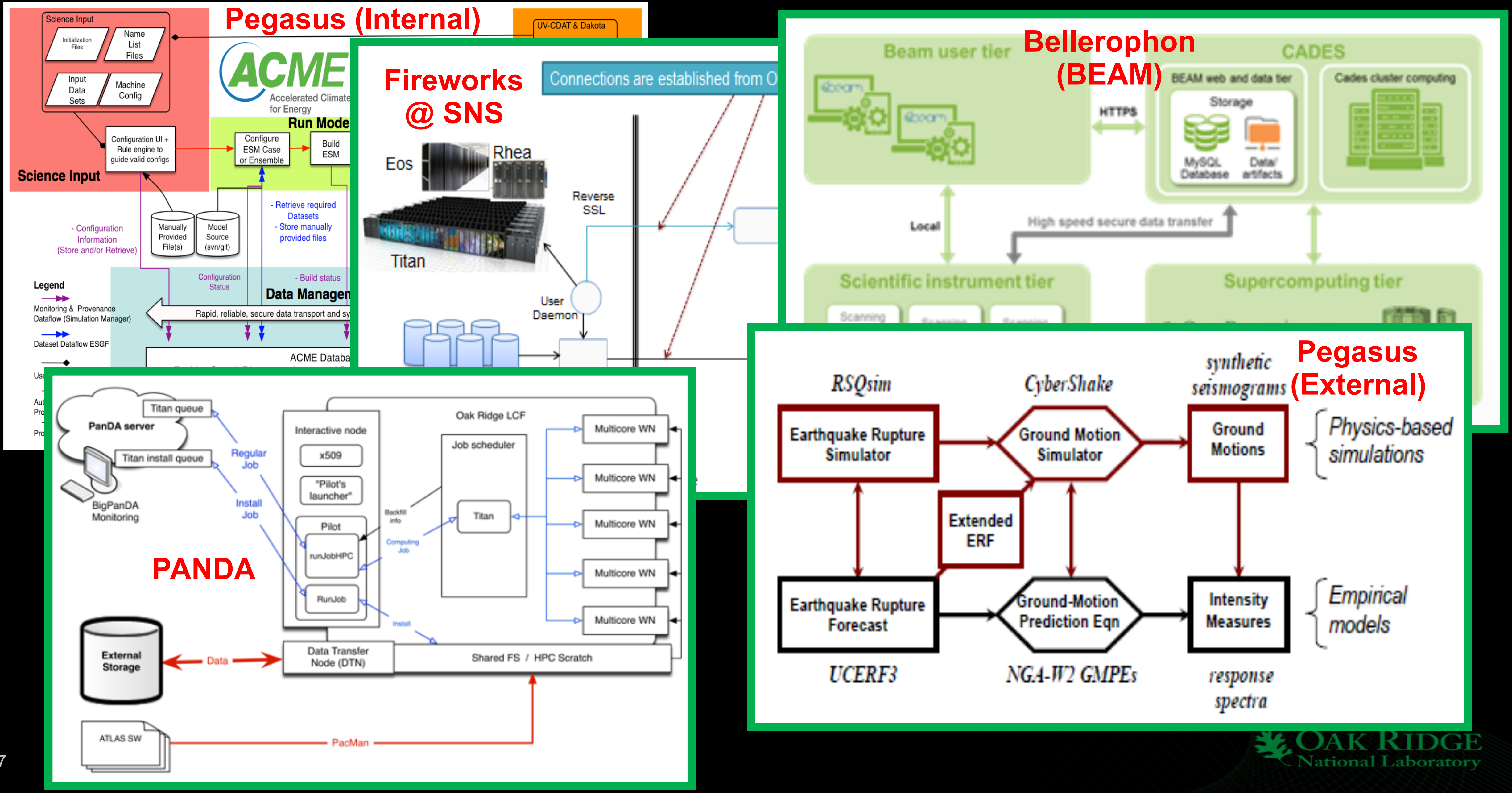
EnSight

VisIt

Power-Wall Visualization

EVEREST@OLCF

What have we started doing ? Workflow Support



What are we learning ?

Work with Byung Park

Taxonomy of Workflows

- Compute-intensive (e.g. SCEC)
- Data-intensive (e.g. ACME)
- Experimental workflows (e.g. GWAS)
- Data Portals (e.g. HACC)

Minimum

- Constant availability and resiliency
- Capability of launching jobs
- Notification of job termination status
- Capability of transferring input/output

Extended

- Real-time Monitoring
- Real-time intervention on jobs
- Overseeing multiple compute/storage resources
- Mini workflow manager
- Suggestions?

Interface

- Flexibility in inserting new jobs
- Seamless integration with the rest of workflow
- Should act as a work component
- Suggestions?

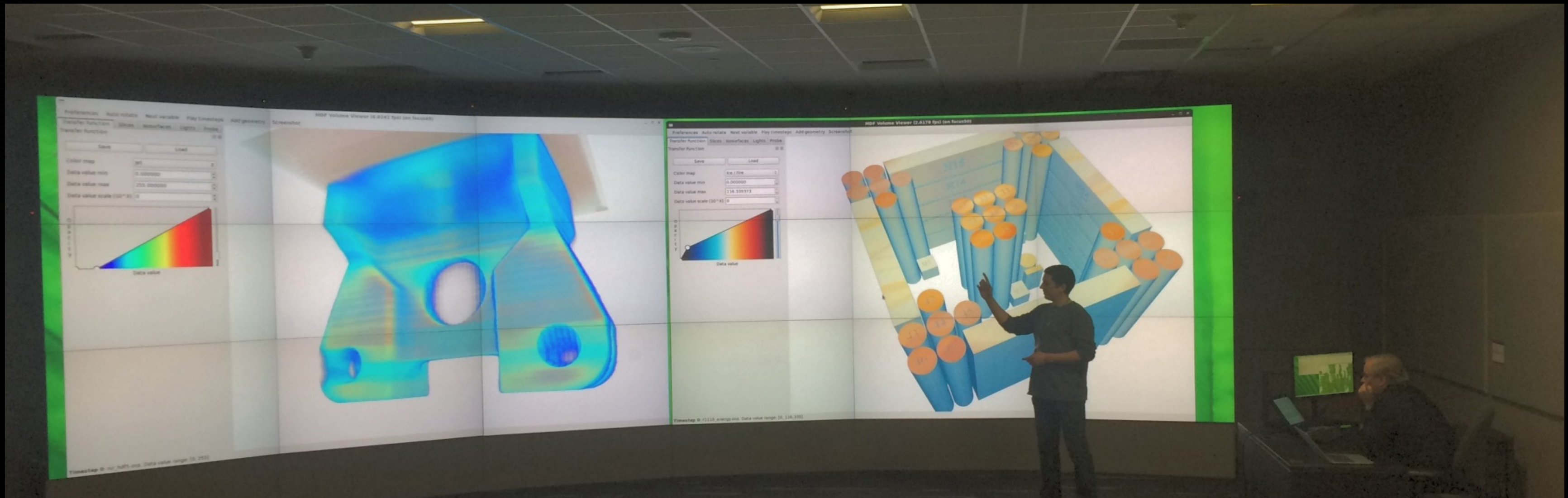
User Stories

Use Case: Big Neuron Hackathon

Ran simulations on Titan and evaluated state-of-the-art neuron reconstruction methods on EVEREST. (Over 100 TBs of image data processed and analyzed in 3 days)

Attendees representing 13 organizations representing North America, Australia and Asia (INCF, OECD) in government and the private sector – Allen Institute for Brain Sciences, George Mason University, etc.

Use Case: Manufacturing Demonstration Facility



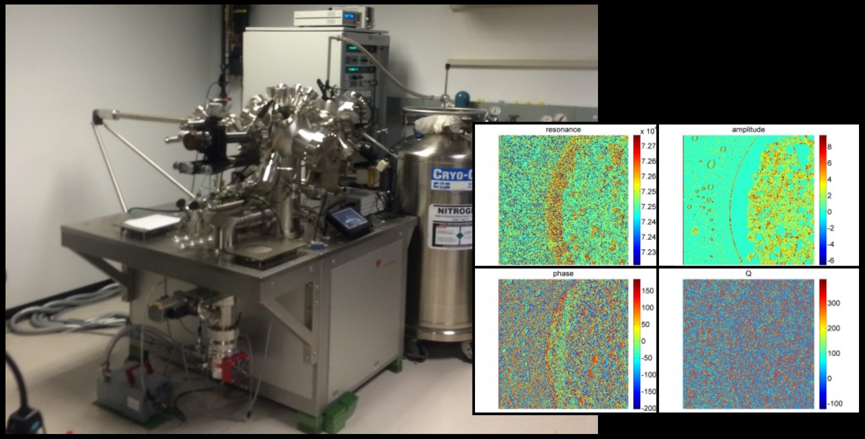
A workflow to predict and interactively visualize artifacts such as porosities and structural deformities based on 3D-printer logs.

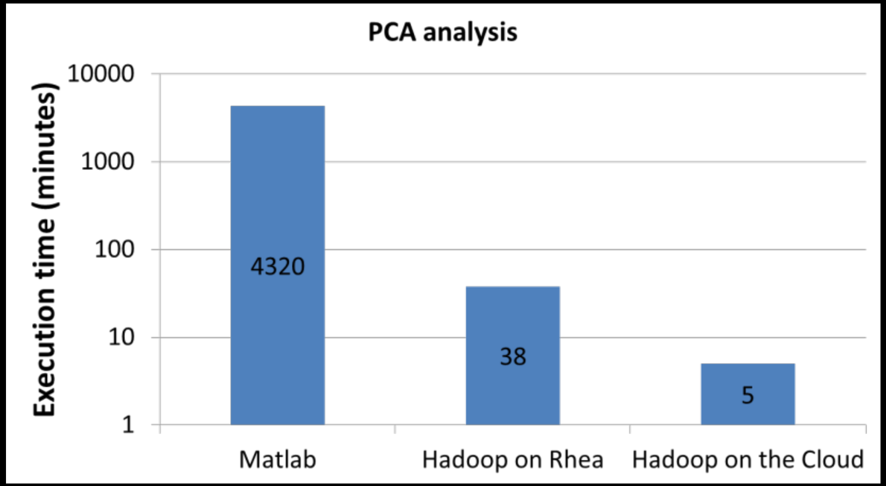
Use Case: Center for Nanophase Materials

How a data scientist can help the domain scientist ?

Scale-up: Instrument captures 1024 by 1024 image at 16000 different spectral bands

Speed-up: Principal component analysis of the image sequence



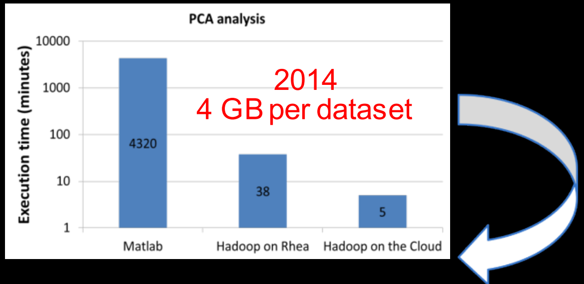


Given a data analysis algorithm of interest to a domain scientist, we can

- recommend optimal ‘analytic’ architecture for speed-up and scale-up.
- quantify performance (cost and latency) trade-offs while using accessible instead of optimal hardware.

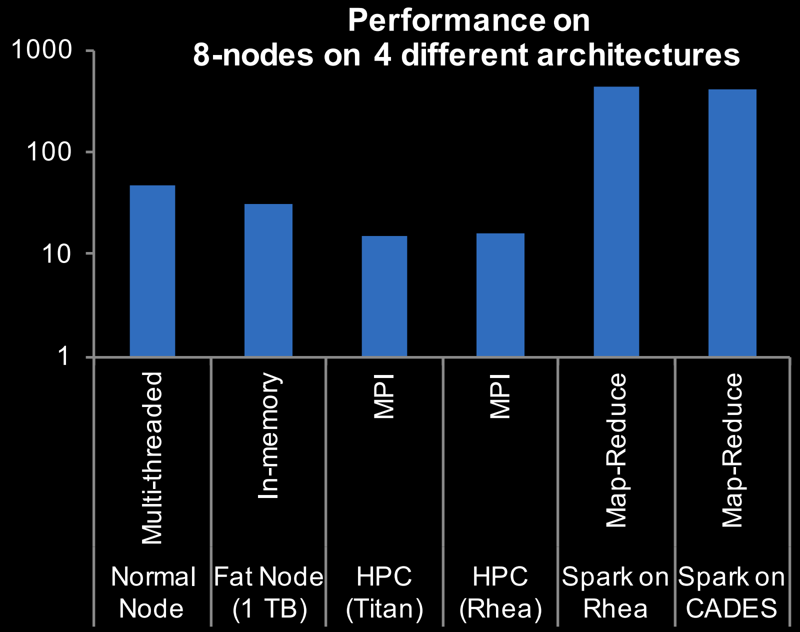
OAK RIDGE National Laboratory

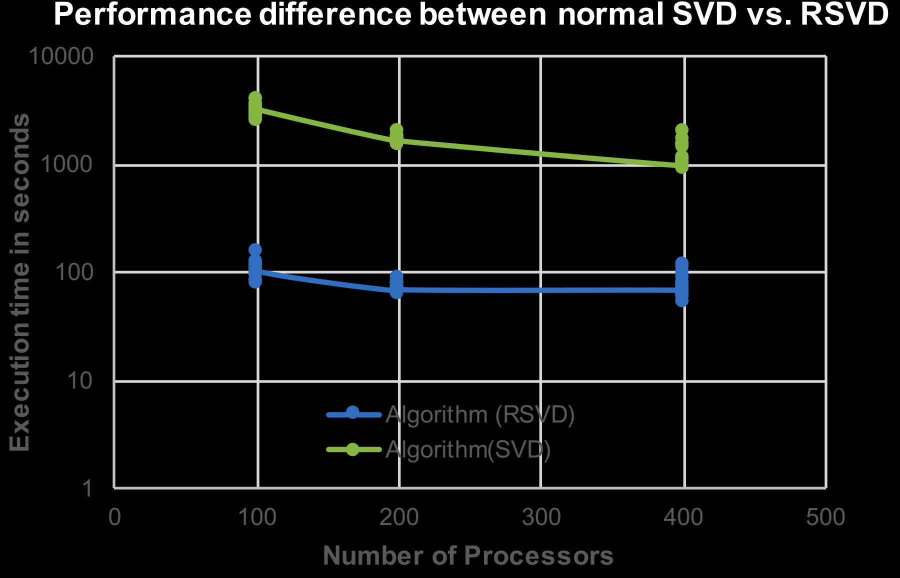
Use Case: Center for Nanophase Materials

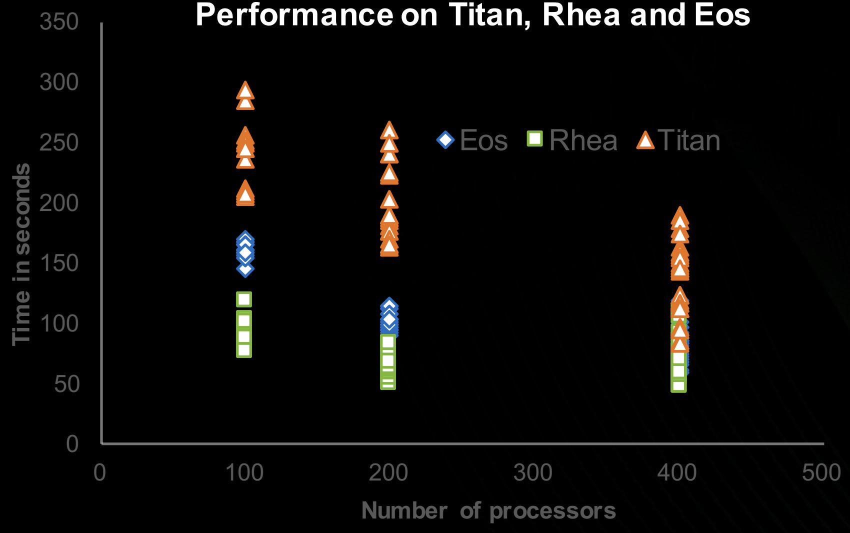


2016: 134 GB per dataset

# of Cores	Speed-up Fat-node	Speed-up Rhea	Execution time (seconds) Fat-node	Execution time (seconds) Rhea
1	1.00x	1.00x	2747	4188
2	1.65x	1.75x	1663	2396
4	2.46x	2.87x	1117	1459
8	2.96x	3.65x	928	1146







OAK RIDGE National Laboratory

Thank You… Questions ?

- How can we help you ? Please chat with our members when you get a chance…..
  - Collect surveys…..

How are we helping our science users ?

Algorithm → Code → Program

- **What hardware to buy/use ?**
  - Investment $
  - Technology
  - Flexibility for growth

- **What is the cost of performance ?**
  - Portability
  - Energy
  - Time-to solution